An investigation into eukaryotic pseudouridine synthases

King, Ross Donald; Lu, Chuan

Published in:
Journal of Bioinformatics and Computational Biology

DOI:
10.1142/S0219720014500152

Publication date:
2014

Citation for published version (APA):
King, R. D., & Lu, C. (2014). An investigation into eukaryotic pseudouridine synthases. *Journal of Bioinformatics and Computational Biology*, *12*(4). https://doi.org/10.1142/S0219720014500152

Journal of Bioinformatics and Computational Biology
Vol. 12, No. 4 (2014) 1450015 (16 pages)

DOI: 10.1142/S0219720014500152

ICP Imperial College Press
www.icpress.co.uk

An investigation into eukaryotic pseudouridine synthases

Ross D. King*,‡ and Chuan Lu†,§

*Manchester Institute of Biotechnology
University of Manchester, 131 Princess Street
Manchester, M1 7DN, UK

†Department of Computer Science
Aberystwyth University, Aberystwyth, UK
‡ross.king@manchester.ac.uk
§cul@aber.ac.uk

Received 9 September 2013
Revised 1 June 2014
Accepted 3 June 2014
Published 1 August 2014

A common post-transcriptional modification of RNA is the conversion of uridine to its isomer pseudouridine. We investigated the biological significance of eukaryotic pseudouridine synthases using the yeast *Saccharomyces cerevisiae*. We conducted a comprehensive statistical analysis on growth data from automated perturbation (gene deletion) experiments, and used bi-logistic curve analysis to characterise the yeast phenotypes. The deletant strains displayed different alteration in growth properties, including in some cases enhanced growth and/or biphasic growth curves not seen in wild-type strains under matched conditions. These results demonstrate that disrupting pseudouridine synthases can have a significant qualitative effect on growth. We further investigated the significance of post-transcriptional pseudouridine modification through investigation of the scientific literature. We found that (1) In *Toxoplasma gondii*, a pseudouridine synthase gene is critical in cellular differentiation between the two asexual forms: Tachyzoites and bradyzoites; (2) Mutation of pseudouridine synthase genes has also been implicated in human diseases (mitochondrial myopathy and sideroblastic anemia (MLASA); dyskeratosis congenita). Taken together, these results are consistent with pseudouridine synthases having a Gene Ontology function of "biological regulation".

Keywords: RNA; *Saccharomyces cerevisiae*; growth analysis; biological regulation.

1. Introduction

A common post-transcriptional modification of RNA is the conversion of uridine to its isomer pseudouridine (Ψ). This isomerisation requires no co-factor(s) and no energy source, and the enzyme catalysing this reaction, pseudouridine synthase, is found in all branches of life.[1] Pseudouridine is present in many types of RNA, and is particularly abundant in small stable RNAs such as transfer RNA (tRNA), small nuclear RNA (snRNA), and in ribosomal RNA (rRNA). Because Ψ has a carbon at

the 5th position of the uracil ring attached to the sugar, rather than the nitrogen at the 1st position, Ψ is potentially more versatile in its hydrogen bonding interactions than uridine, strengthening the stems of RNA secondary structures, and stabilising base stacking in loops. The post-transcriptional modification of uridine to pseudouridine involves numerous molecular agents: Pseudouridine synthases and H/ACA box small nucleolar RNAs (snoRNAs) with associated proteins. Eukaryotic pseudouridine synthases operate in two major ways — either as distinct enzymes without any additional co-factors, or as directed by the guide snoRNAs.[2] Deletion of the guide RNA directed PUS enzymes (CBF5) typically cause lethality or severe phenotypes since they are responsible for a large number (tens to hundreds) of ribosomal psi modifications.

In this paper we investigate the biological significance of pseudouridine synthases in eukaryotes, with a focus on the yeast *Saccharomyces cerevisiae*. Currently in yeast, most experimental evidence indicates that the primary target of pseudouridine synthases are tRNAs (Table 1). However, pseudouridine post-transcriptional modification is also common in rRNA, where it plays an important role (Table 1).

2. Methods

2.1. *Pseudouridine synthase genes in yeast*

We focused our experimental investigation on pseudouridine post-transcriptional modification in the yeast *S. cerevisiae* as it is genetically tractable and the organism of choice for studying the eukaryotic cell. We identified ten pseudouridine synthase genes (Table 1). All pseudouridine synthases are believed to share the same fold of their catalytic domain(s) and uracil-binding site, and are believed to descend from a common molecular ancestor.[3,4] A reasonable amount is known about the molecular functions of the *S. cerevisiae* pseudouridine synthases[3,5] (see Table 1).

2.2. *Perturbation experiments*

To investigate pseudouridine synthase function we executed perturbation growth experiments using the EUROSCARF strain BY4741 (MATa haploids, HIS3Δ, LEU2Δ, MET15Δ, URA3Δ) as the wild type (WT). To ensure that these nutritional markers did not qualitatively change the behaviour of the WT we also investigated as a control the strain YSBN5 (MATa, FY3 ho::Ble). We employed the single deletant strains from EUROSCARF collection of MATa haploids with WT BY4741. From the Saccharomyces Genome Database (SGD)[3] we selected the deletant strains corresponding to the deletion of the following genes: PUS1, PUS2, PUS3, PUS4, PUS5, PUS6, PUS7, PUS8, PUS9, CBF5. Two of these strains PUS8Δ and CBF5Δ are not viable haploids: The PUS8 gene product is bifunctional, catalysing the third step in riboflavin biosynthesis as well as isomerisation of U to pseudouridine, and it has been shown that only the domain involved in riboflavin synthesis is required for growth[6]; and the gene CBF5 has a function distinct from the other

Table 1. The existing functional annotation of the pseudouridine synthase genes, taken from the *Saccharomyces* Genome Database.[3]

Gene	ORF id	Current annotated molecular function
PUS1	YPL212C	tRNA: Pseudouridine synthase, introduces pseudouridines at positions 26–28, 34–36, 65, and 67 of tRNA; nuclear protein that appears to be involved in tRNA export; also acts on U2 snRNA.
PUS2	YGL063W	Mitochondrial tRNA: Pseudouridine synthase; acts at positions 27 and 28, but not at position 72; efficiently and rapidly targeted to mitochondria, specifically dedicated to mitochondrial tRNA post-transcriptional modification.
PUS3 (DEG1, HRM3)	YFL001W	tRNA: Pseudouridine synthase, introduces pseudouridines at position 38 or 39 in tRNA, important for maintenance of translation efficiency and normal cell growth, localises to both the nucleus and cytoplasm.
PUS4	YNL292W	Pseudouridine synthase, catalyses only the formation of pseudouridine-55 (Psi55), a highly conserved tRNA post-transcriptional modification, in mitochondrial and cytoplasmic tRNAs.
PUS5	YLR165C	Pseudouridine synthase, catalyses only the formation of pseudouridine (Psi)-2819 in mitochondrial 21S rRNA.
PUS6	YGR169C	tRNA: Pseudouridine synthase, catalyses the conversion of uridine to pseudouridine at position 31 in cytoplasmic and mitochondrial tRNAs.
PUS7	YOR243C	Pseudouridine synthase, catalyses pseudouridylation at position 35 in U2 snRNA, position 50 in 5S rRNA, position 13 in cytoplasmic tRNAs, and position 35 in pre-tRNA(Tyr). Stress response.
PUS8 (RIB2)	YOL066C	Bifunctional enzyme with DRAP deaminase and tRNA: Pseudouridine synthase activity; the deaminase catalyses the third step in riboflavin biosynthesis and the synthase catalyses formation of pseudouridine at position 32 in cytoplasmic tRNAs.
PUS9	YDL036C	Mitochondrial tRNA: Pseudouridine synthase, catalyses the formation of pseudouridine at position 32 in mitochondrial tRNAs; contains an N-terminal mitochondrial targeting sequence.
CBF5	YLR175W	Pseudouridine synthase catalytic subunit of box H/ACA small nucleolar ribonucleoprotein particles (snoRNPs), acts on both large and small rRNAs and on snRNA U2.

pseudouridine synthases (it is a subunit of box H/ACA small nucleolar ribonucleoprotein particles (snoRNPs)),[7] and this may be the reason for lethality. So far the other eight genes have only been annotated to pseudouridine synthase activities. The deletion strain COX6Δ was also used in these experiments as a negative control for growth with deficient mitochondria.

2.3. *Growth medium*

The basic growth medium used in the *S. cerevisiae* growth experiments was a standard defined ("minimal") medium (containing DIFCO yeast nitrogen base without amino acids, dextro and some supplements), with glucose (1.2% w/v) as the main carbon source, supplemented with histidine, leucine, methionine, and uracil. The rich medium used was yeast extract-peptone-dextrose (YPD), which is a complete medium for yeast growth.

2.4. *Automated growth assays*

To investigate the phenotypes of the deletant strains we utilised our established robust laboratory automation protocol for a high-throughput assay for growth curves using our "Robot Scientist" Adam.[8,9] This automated protocol enables Adam to accurately measure and compare yeast growth curves. Adam's basic operations are selection of specified yeast strains from a library held in a freezer, inoculation of these strains into microtiter plate wells containing rich medium, measurement of growth curves on rich medium, harvesting of a defined quantity of cells from each well, inoculation of these cells into wells containing defined media, and measurement of growth curves on the specified media. The deletant strains were first grown in rich media for around 24 h at a temperature of 30°C before they were transferred to the test media. OD readings were then recorded regularly for 72 h. Each strain was grown in a 96 well plate with controls (and possibly other strains of interest). In each plate, there are around 24 replicates for the same strain, and each strain have been grown in least two different plates and sometimes in different batches (with different time periods). An OD measurement is made every ~ 20 min in the defined medium (~ 40 min in rich). To improve the accuracy and reliability of its readings, Adam used shaking incubators, and a Variomag Teleshake to agitate plates before readings. The deletant strains were grown in 96-well plate with the WT as positive controls. Growth was not anaerobic (no sealing of plates); otherwise no diauxic shift could have been observed.

The OD observations for each well were processed by extracting a list of growth curve parameters from the curves smoothed by spline fitting.[8] The three main parameters we used for this study are global maximum OD, maximum growth rate, and the lag time for reaching the maximum exponential growth. The growth curve parameters were stored in a relational database for statistical analysis, along with their corresponding metadata including strains, media, well location, plate id, etc. For this analysis we use the relative global maximum OD (to the WT) throughout the measurement cycle as the main estimate of the general growth fitness.

Compared to previous yeast growth experiments involving pseudouridine synthase gene deletant strains, our experiments have the following advantages: (1) The growth curves can capture some of the dynamic phenotypic characters (instead of just growth fitness measured from one feature such as colony size); (2) cell growth is monitored for 72 h, which enables us to assess the phenotypes after more generations; (3) A relatively large number of replicates within and cross plates with controls may allow us to draw conclusions about the phenotypes with more confidence.

2.5. *Statistical data analysis*

The data consisting of the curve parameters and meta-information was processed for quality control and outlier detection. Outliers were removed from further analysis; the main criteria for outlier data points were:

(1) Wells for which the global maximum OD measure in pre-growth had not reached a required threshold: This is an indication of abnormality in

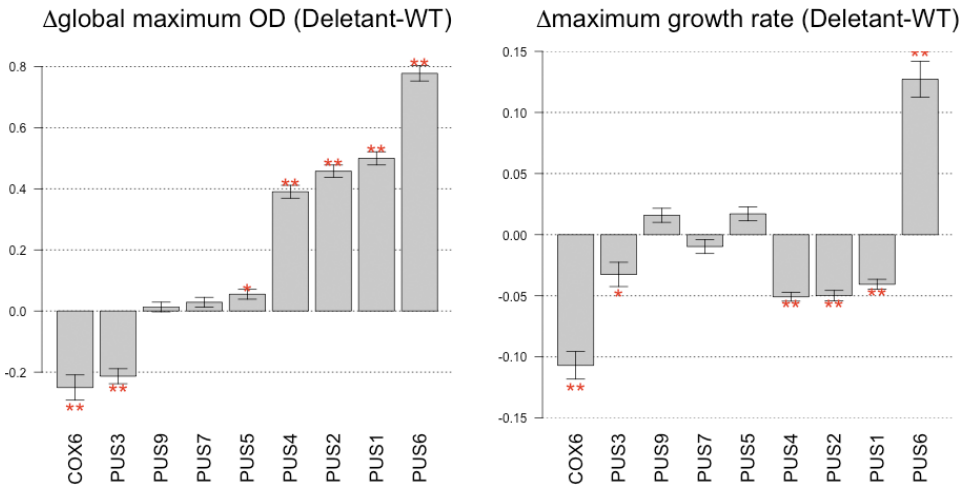
cell growth due to various experimental factors (such as defects in inoculation).
(2) Wells with Mahalanobis distance outside the 75% confidence intervals based on replicate wells of the same strain/media combination in the same plate. The Mahalanobis distance was calculated using the three main curve parameters.

After removing the outlying points, basic descriptive statistics were calculated, and statistical tests conducted to compare the key curve parameters including the maximum growth rate, the global maximum OD, and the lag time for reaching exponential growth. The relative growth fitness estimate (the gross difference between the strain and the median of the WT strain of the same plate and the same medium) was used to test the statistical difference between the WT and the deletant strains using t-tests and Wilcoxon rank sum tests.

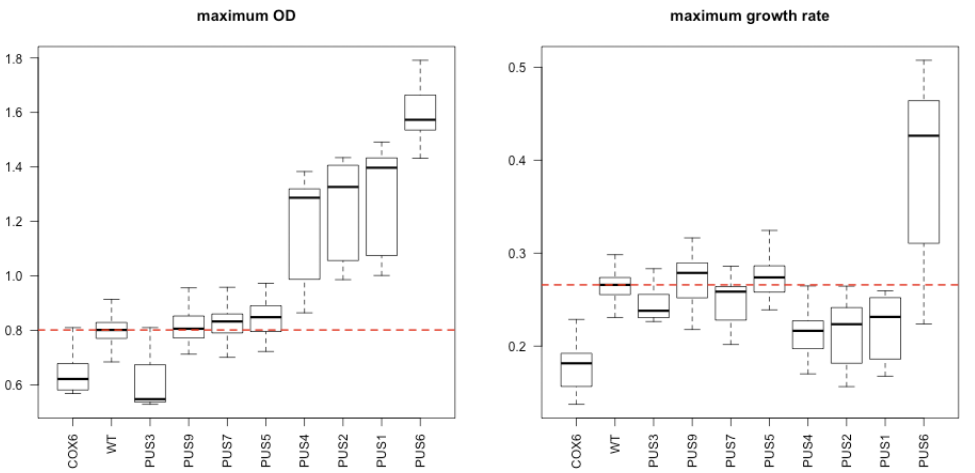
For each strain, a linear mixed effect model was used to test the fixed effect of the gene deletion over the WT strain, while taking into account the random effect for plates[10] (see Fig. 1). The mixed model for cross-plate normalisation of the growth fitness parameter is $y_{s,p} = \mu + \beta * s + u * p + e$ where $y_{s,p}$ represents the growth fitness estimate, i.e. the global maximum OD for strain s in plate p, μ is the model intercept representing the global mean of WT, β as the fixed effect for the strain s (i.e. the difference between strain and WT normalised across plates), a random effect term $u \sim N(0, \sigma_u)$, and a random error term $e \sim N(0, \sigma_e)$. The coefficient and significance of the effect for strains (with a Bonferroni correction for multiple tests) was assessed and reported.

In batch culture experiments it is sometimes unsatisfactory to describe growth phenotypes using a single value estimate for maximum growth rate or maximum OD, especially in case of biphasic growth. Therefore, the following procedure was followed to identify biphasic growth. First, we applied both single-logistic and bi-logistic curve fitting[11] to OD readings of every well, and compared the goodness-of-fit between the two. The bi-logistic curve fitting was carried out by nonlinear least squares optimisation on the curves after outlier removal. A likelihood ratio test was employed to check if a bi-logistic model fits better than a single logistic model. Furthermore, biphasic growth within individual wells was identified based on the parameter estimates of the bi-logistic model about the mixture of two growth pulses. A curve was then identified as exhibiting biphasic growth behaviour when the following criteria were satisfied:

(1) The bi-logistic curve fits significantly better than a standard single logistic curve (with p-value < 0.01 for a likelihood ratio test) and should achieve a $R^2 > 0.95$.
(2) The type of the bi-logistic curve should be sequential,[11] describing a curve with two almost non-overlapping logistic growth pulses. The second pulse does not start growing until the first pulse has reached 90% of saturation of its capacity. The shape characterises a system that pauses between growth phases. In this study, taking into account the variability from the biological systems and the



(a)



(b)

Fig. 1. (a) The effect of gene deletion over the WT strain in minimal media. The effect was estimated using the coefficients obtained from the linear mixed models with strain as the fixed effect and the plate as the random effect. The positive effect indicates the higher global maximum OD (left panel) or higher maximum growth rate (right panel) compared to the WT and vice versa. The error bars indicate the standard errors for the estimated effect. The number of stars represents the significance level of the deletion effect: Two stars indicating level of 0.01; one star level of 0.05; no star not significant. The significance levels were estimated based on the p-values for the fixed effect coefficients with Bonfferoni correction. The results of COX6Δ are reported as a negative control for growth with deficient mitochondria, and (b) boxplots of the global maximum OD and maximum growth rate extracted from the growth curve for minimal medium. The dashed line indicates the median for the WT strain. The growth fitness data shown here have been normalized such that the median value for the WT strain will be the same for every plate, while keeping the original relative difference between the WT and the other strains.

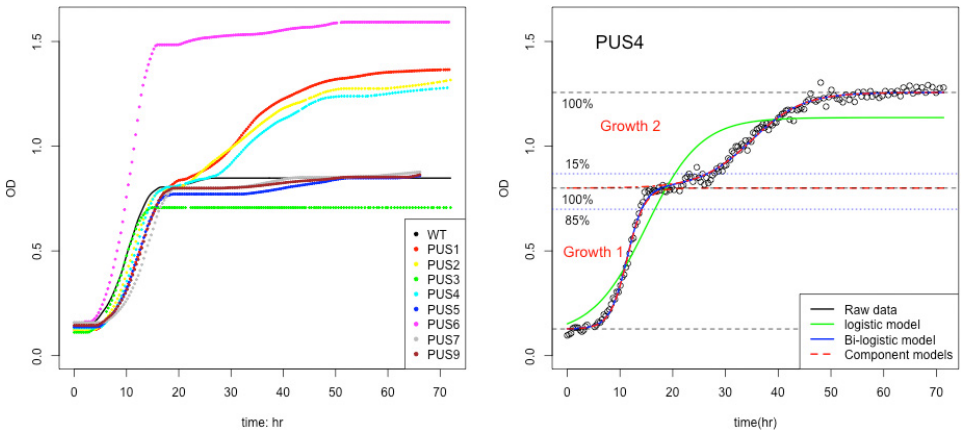


Fig. 2. Typical growth curves for the eight deletant strains compared to the WT growing in minimal medium. The left panel shows the growth curves (smoothed by the LOWESS method for illustrative purpose), which were randomly selected from the dataset and have global maximum OD and maximum growth rate close to the median for that strain. The right panel gives the example of bi-logistic curve fitting (the blue curve) given the OD readings of a growth curve (the black circles) compared to the single logistic curve fitting (the green curve). The two component models representing different growth pulses were indicated in red dashed lines. The bottom dashed line corresponds to the baseline of the bi-logistic model; the two dashed lines above the baseline indicate the levels when the first growth pulse reaches its 85% and 100% capacity. The 15% and 100% capacity levels for the second growth pulse were indicated in the two dashed lines starting from the 100% capacity level of the first growth. Here, the well-fitted bi-logistic model indicates a biphasic growth, for which the second growth pulse does not reach its 15% capacity before the first growth reaches its 85% capacity.

measurement system, the non-overlapping criterion for sequential growth was slightly relaxed. We assume that both growth processes for a sequential logistic curve with a biphasic growth could be overlapping to a certain degree: The second growth pulse reaches 15% of its capacity only after the first growth pulse reaches its 85% capacity (see Fig. 2).

(3) To ensure the significance of the biphasic growth characteristics, we assumed that the capacity (i.e. the maximum added biomass measured in OD) of the second growth process should be greater than 0.1; we also excluded growth curves that did not reach a stationary phase (with first derivatives < 0.02) at the end of the 72 h. measurement period.

To identify a phenotype having biphasic growth behaviour Chi-squared tests were conducted to check whether the deletant strain has a higher proportion of curves showing biphasic growth behaviour compared to the WT. About 9% of WT growth curves exhibited biphasic growth behaviour (after excluding outlying data) in this study, the significance cut-off p-value is set to be 0.005 — taking into account multiple comparisons.

All the statistical analyses were run using *R* and are available on request.

3. Results

3.1. *Growth curve analysis*

We compared the growth of all the viable yeast deletant pseudouridine synthase strains in the standard defined growth medium. This resulted in the discovery of both qualitative and quantitative differences from the WT (BY4741) (Tables 2 and 3; Figs. 1 and 2).

3.2. *Biphasic growth curves*

Under our standard growth conditions the WT does not exhibit a biphasic growth curve (typical of a diauxic shift). Instead its growth curve resembles a logistic curve with a sharp cut-off (Fig. 2). To test whether the lack of an observed biphasic growth was caused by the nutritional markers in BY4741 we compared the strain's growth with that of the strain YSBN5 (with no nutritional markers), but we found no qualitative difference using our standard growth medium and our automated growth assays.

Compared to the WT, the deletant strains PUS1Δ, PUS2Δ, PUS4Δ, exhibit qualitatively different growth (Fig. 2). Their biphasic growth is demonstrated by the large R^2 (usually > 0.95) for the bi-logistic curve fitting, and the significant

Table 2(a). Descriptive statistics on the growth fitness. Two growth parameters relative to the WT median of each plate have been used to measure the growth fitness: Global maximum OD (global_max_od) and maximum growth rate (linear_slope). The unadjusted p-values for comparison of the group means for deletant strains and the WT strain based on two-sample t-tests and Wilcoxon rank sum tests have also been reported. The outlying data have been removed prior to further statistical analysis.

Par	Strain	nRep	nPlates	median	mad	mean	sd	p_t test	p_wilcox
Global_max_od	COX6	14	2	−0.180	0.070	−0.248	0.258	0.0044	4.0E−08
	WT	207	13	0.000	0.043	−0.002	0.078	1.0000	1
	PUS9	36	4	0.005	0.056	0.007	0.069	0.4523	0.547
	PUS5	35	4	0.045	0.077	0.050	0.069	0.0006	0.002
	PUS3	14	3	−0.254	0.058	−0.219	0.165	0.0003	5.3E−07
	PUS7	35	4	0.031	0.055	0.023	0.067	0.0689	0.082
	PUS4	27	3	0.485	0.070	0.386	0.177	9.0E−12	1.4E−13
	PUS2	28	3	0.525	0.136	0.455	0.166	1.4E−14	5.4E−14
	PUS1	27	3	0.595	0.080	0.496	0.182	5.0E−14	1.1E−13
	PUS6	38	4	0.771	0.088	0.791	0.083	1.3E−37	2.9E−20
Linear_slope	COX6	14	2	−0.084	0.028	−0.105	0.072	0.0001	1.6E−10
	WT	207	13	0.000	0.014	−0.003	0.022	1	1
	PUS9	36	4	0.013	0.021	0.006	0.027	0.0233	0.017
	PUS5	35	4	0.008	0.020	0.008	0.022	0.0055	0.023
	PUS3	14	3	−0.028	0.014	−0.035	0.073	0.1094	0.002
	PUS7	35	4	−0.007	0.016	−0.017	0.021	0.0501	0.014
	PUS4	27	3	−0.049	0.020	−0.052	0.026	6.1E−11	2.5E−12
	PUS2	28	3	−0.042	0.048	−0.051	0.035	2.6E−08	2.0E−11
	PUS1	27	3	−0.035	0.036	−0.042	0.033	8.9E−07	5.2E−11
	PUS6	38	4	0.160	0.084	0.127	0.085	3.6E−11	4.5E−10

Table 2(b). Results of linear mixed model analysis on growth fitness for each deletant mutant (MT) strain versus WT strain in minimal media. Two growth parameters including global maxium OD and maximum growth rate have been used as growth fitness measures. For each growth parameter, a mixed model was built for each deletion strain taking strain as the fixed effect the plate as the random effect. The coefficients for the intercept correspond to the medians of the WT across the plates containing the deletion strains under tests. The coefficients for the strain and the corresponding p-values indicate the fixed effect of deletion strain (the change in growth fitness relative to the WT) in minimal media. The p-values reported here have not been corrected for multiple comparisons and were obtained from individual linear mixed models.

	Strain	nRep (MT)	nRep (WT)	Coef. intercept	Coef. strain	p val. intercept	p val. strain	SE. intercept	SE. strain
Global_max_od	COX6	14	39	0.860	−0.250	3.32E−31	1.8E−07	0.032	0.041
	PUS1	27	60	0.902	0.500	7.95E−16	5.0E−38	0.091	0.022
	PUS2	28	60	0.903	0.460	6.75E−15	1.1E−37	0.095	0.020
	PUS3	14	60	0.904	−0.213	3.40E−10	1.0E−13	0.124	0.025
	PUS4	27	60	0.903	0.390	1.34E−15	3.6E−31	0.092	0.021
	PUS5	35	73	0.790	0.055	2.21E−68	0.0009	0.018	0.016
	PUS6	38	35	0.761	0.778	4.73E−38	9.1E−42	0.028	0.025
	PUS7	35	73	0.791	0.029	1.79E−68	0.075	0.018	0.016
	PUS9	36	73	0.790	0.013	8.62E−70	0.408	0.018	0.016
linear_slope	COX6	14	39	0.257	−0.107	7.64E−42	8.4E−13	0.006	0.011
	PUS1	27	60	0.287	−0.041	2.52E−43	9.8E−16	0.011	0.004
	PUS2	28	60	0.287	−0.050	1.87E−42	1.9E−19	0.011	0.004
	PUS3	14	60	0.288	−0.033	5.48E−25	0.002	0.018	0.010
	PUS4	27	60	0.288	−0.051	2.10E−36	4.9E−24	0.013	0.004
	PUS5	35	73	0.267	0.017	7.58E−96	0.004	0.003	0.006
	PUS6	38	35	0.238	0.127	4.27E−32	1.3E−12	0.011	0.015
	PUS7	35	73	0.267	−0.010	3.93E−94	0.091	0.003	0.006
	PUS9	36	73	0.267	0.016	1.22E−95	0.007	0.003	0.006

improvement in goodness-of-fit according to the likelihood ratio tests compared to the standard logistic curve fitting. It is natural to infer that the biphasic growth curves observed in PUS1Δ, PUS2Δ, PUS4Δ are caused by anomalous diauxic shifts. Such shifts involve the coordinated expression of a large number of genes.[12,13]

Table 3(a). Comparison of the proportions of wells that have distinctive biphasic growth behaviour between the deletant strains and the WT strain. Outlying data have been removed prior to this comparison. Chi-squared tests have been used to check the statistical significance in the difference between the two proportions.

Strain	#Biphasic	#Replicates	Rate_biphasic (%)	Chisqtest_p value
PUS4	26	27	96.3	5.45E−26
PUS2	20	28	71.4	9.06E−16
PUS1	8	27	29.6	0.005
PUS9	7	36	19.4	0.122
PUS5	6	35	17.1	0.258
PUS6	6	38	15.8	0.344
PUS7	4	35	11.4	0.914
PUS3	1	14	7.1	1
COX6	0	14	0.00	0.488
WT (control)	19	207	9.2	NA

Table 3(b). The comparison of observed Max OD (OD) and growth rate (R) relative to that of the WT: ++ large positive difference from WT; + positive difference from WT; = no significant difference from WT; −− large negative difference from WT; − negative difference from WT.

	Growth		
Gene	OD	R	Biphasic growth
WT	na	na	×
PUS1	++	−−	✓
PUS2	++	−−	✓
PUS3	−−	−	×
PUS4	++	−−	✓
PUS5	+	=	×
PUS6	++	++	×
PUS7	=	=	×
PUS9	=	=	×

3.3. *Quantitative analysis of growth curves*

We also carried out a quantitative comparison of the deletant strain growth curves with the WT (Tables 2(a) and 2(b) and Figs. 1(a) and 1(b)). This found that the biphasic growth of strains PUS1Δ, PUS2Δ, PUS4Δ gave them a much greater maximum growth yield: A maximum OD reading more than 50% that of the WT. The growth yield of PUS5Δ was also slightly greater than the WT, and that of PUS6Δ is significantly greater. In addition, Table 3(a) shows that the rate of significant biphasic growth for PUS5Δ and PUS6Δ is 17% and 16% respectively, in contrast to 9% for WT. Although these differences are not significant in a Chi-squared test, the balance of evidence favours a difference, and it is possible that these deletant strains actually involve in faster metabolic reprogramming processes that are not easy to detect in our experiments — in which the cell growth was monitored by OD measures every 20 min. If the cells stay in biphasic shifts for less than 20 min, then such subtle change in cell growth might not be observable.

4. Discussion

4.1. *Possible mechanisms for the anomalous growth observed*

It is natural to speculate on the mechanism that enables the deletant stains PUS1Δ, PUS2Δ, PUS4Δ to achieve much greater yield, and in the case of PUS6Δ both greater yield and growth rate. There would seem to be two most likely explanations: The deletants use some resource in the growth medium (probably N or P) more efficiently enabling the remaining C resource to be utilised, or the deletants have lost the ability to induce cell cycle arrest under growth conditions, the WT has evolved to

consider unfavourable. Either explanation would seem to require some sort of change in biological regulation.

4.2. *Multiple functions for "pseudouridine synthase" genes*

It is logically possible to argue that the observed phenotypes with "pseudouridine synthase" gene deletion are not caused by their pseudouridine synthase function; but rather by some additional functions that they may have. Indeed, we have used this argument ourselves to explain the lethality of the pseudouridine synthase gene PUS8Δ, which is bi-functional and also catalyses the third step in riboflavin biosynthesis. However, while it is possible that some of the evidence we have presented is the result of multi-functions, it would seem improbable that multiple functions is the explanation for all the evidence for a role of pseudouridine synthases in biological regulation.

4.3. *Evidence from the scientific literature on the role of pseudouridine synthases*

To further investigate the role of pseudouridine post-transcriptional modification by pseudouridine synthases we searched the literature for information on the phenotypes of organisms with altered pseudouridine synthase genes. We found the following (Table 4):

(1) In *S. cerevisiae*, PUS7 is involved in conditional pseudouridine post-transcriptional modification in response to environmental stress (heat-shock and nutrient deprivation).[14] This result was the first direct demonstration that pseudouridine post-transcriptional modification is inducible and involved in regulation.
(2) In *Toxoplasma gondii*, the homolog to PUS9 is critical in cellular differentiation. *T. gondii* is an important human protozoan parasite. A key part of *T. gondii's* ability to establish a lifelong chronic infection and evade the immune system, is its developmental switch between the two asexual forms: tachyzoites and bradyzoites. This implies that the *T. gondii* homolog of PUS9 is central to this developmental switch.[15]
(3) Mutations in the human homolog to PUS1 cause the disorders mitochondrial myopathy and sideroblastic anemia (MLASA).[16] MLSA is a rare autosomal recessive oxidative phosphorylation disorder specific to skeletal muscle and bone

Table 4. Notable disease phenotypes of pseudouridine synthase genes.

Gene	Disease phenotype
PUS7	Stress response in *S. cerevisiae*
PUS1	Mitochondrial myopathy and sideroblastic anemia
PUS9	*T. gondii* developmental differentiation
CBF5	Dyskeratosis congenita

Table 5. Summary of evidence for the role of pseudouridine synthase genes in cellular information processing from our growth experiments and the scientific literature: ✓ evidence for, × evidence against, — no evidence.

Gene	Growth	Disease
PUS1	✓	✓
PUS2	✓	—
PUS3	×	—
PUS4	✓	—
PUS5	?	—
PUS6	✓	—
PUS7	✓	—
PUS8	NA	—
PUS9	×	✓
CBF5	NA	✓

marrow. The growth phenotype we have identified for PUS1Δ is consistent with the disease phenotype in humans, as MLSA is primarily a disease of mitochondria.[17]

(4) Mutations in the human homolog to CBF5 (dyskerin, DCK1) cause the X-linked form of the disorder dyskeratosis congenita (DKC).[17] This is a rare progressive congenital disease that resembles premature ageing, and is similar to progeria. Most DKC1 mutations lie in the PUA domain but mutations in the catalytic domain are associated with a severe variant known as Hoyeraal–Hreidarsson syndrome.[18] Mutated mutant mice show decreased pseuduridine content in their rRNAs, and specific impairments in transcription.[18] Like MLSA, this syndrome involves damage to bone marrow and anemia. The deletion of CBF5 is lethal, however, the molecular function CBF5 is similar to that of dyskerin.[17]

Table 5 shows a summary of the investigation of the function of pseudouridine synthases in *S. cerevisiae*.

4.4. *Possible molecular mechanisms*

What are molecular mechanisms resulting from the perturbations of pseudouridine synthases that cause the phenotypes described above? There are a number of possible answers — not necessarily mutually exclusive — depending on the type(s) of RNA modified. In yeast, most experimental evidence (with the exception of CBF5) suggests that pseudouridine synthases predominantly act on tRNAs. The natural hypothesis would then be that the post-transcriptional modification of tRNA in some way changes their interpretation by the cell, either by changing their transport, or by changing the translation process. However, despite the current evidence that tRNAs are the target for pseudouridine synthases, *a priori* they are not the most natural

target to cause the phenotypes observed, which suggests that indirect effectors are involved.

It is interesting that mouse PUS1 (its closest yeast homolog is PUS1) can bind with nuclear receptors (NRs) and function as a co-activator through pseudouridylation of an RNA co-activator called steroid receptor RNA activator (SRA).[19] As there are no NRs in yeast, the same mechanism cannot be used. However, yeast does seem to have a closely related set of transcription factors.[20] So it is possible that pseudouridine synthases act on these transcription factors to switch metabolism.

The role of pseudouridine post-transcriptional modification is important in the ribosome: Pseudouridine formation is targeted only to residues that compose the mature ribosome; pseudouridine residues are found disproportionally in important regions of the ribosome; pseudouridine bases are conserved from bacteria to mammals; specific roles for pseudouridine post-transcriptional modification in ribosome–ligand interactions are conserved in yeast, mouse, and humans; the number of pseudouridine bases in the ribosome increases dramatically from eubacteria to eukaryotes; and changes in some specific pseudouridines are non-lethal.[18] Therefore, perturbations of pseudouridine synthases may affect growth through an indirect effect on ribosomes.

4.5. *Future work*

To further investigate pseudouridine synthases, more experiments are required both to ensure that the phenotypes observed when they are perturbed are caused by pseudouridine post-transcriptional modification, and to determine the molecular mechanisms that cause the phenotypes. For example, it may be possible to use protein engineering to produce proteins that retain/lose the pseudouridine synthase function while loosing/retaining the other functions. A precedent for this is the removal of pseudouridine synthase function in CBF5.[21]

4.6. *Biological regulation using pseudouridine post-transcriptional modification?*

Cells use modification of DNA and protein for biological regulation. The best understood epigenetic mechanism is the methylation and de-methylation of DNA — mostly at CpG sites, with cytosine being converted to 5-methylcytosine. In methylation, at the physical level, the local structural and chemical properties of a DNA molecule are changed, while at the abstract level the DNA is labelled with information. Similarly, in protein signalling, information is processed by the post-translational modifications of protein structures, with the most common modification being the phosphorylation and de-phosphorylation of selected residues. In phosphorylation, at the physical level the structural and chemical properties of a protein are changed, while at the abstract level the protein is labelled with information.

In contrast to the well-recognised importance of biological regulation by the modifications of DNA and proteins, the possible biological regulation role of post-transcriptional labelling of RNA has been little studied. Despite this neglect, it would seem *a priori* likely given biological systems propensity to use all available means to achieve their ends, that post-transcriptional modification of RNA is used for some form of biological regulation. The main argument against this is that pseudouridine post-transcriptional modification is generally considered to be irreversible. However, irreversibility does not preclude the use of pseudouridine modification in regulation, nor is it impossible that pseudouridine post-transcriptional modification is reversible. Of course just because some pseudouridine post-transcriptional modification are used for regulation does not imply that all such modifications, or even a majority of them are.

The generally accepted authority for describing the function of genes is Gene Ontology (GO).[22] In GO, there is an important distinction between a gene product's "molecular function" and the "biological process" it participates in. Currently (11.12.12) in GO, the biological process "biological regulation" is the main term used for regulation in GO: For example, it is a parent of "protein kinase binding" and "DNA methyltransferases". In GO, the molecular function of the yeast pseudouridine synthases are annotated as "pseudouridine synthase activity" which is an "isomerase activity". The biological process that pseudouridine synthases are involved in is "pseudouridine synthesis" which is a "RNA modification". It is notable that the biological process "biological regulation" is not currently a parent of "pseudouridine synthase activity". Given the evidence presented above, we argue that GO needs to be modified to make "biological regulation" a parent of "pseudouridine synthase activity".

Acknowledgments

We wish to thank Mr. Ron Pateman and Dr. Wayne Aubrey for helping with the growth experiments. This work was supported by grant BB/F008228/1 from the UK Biotechnology & Biological Sciences Research Council.

References

1. Ericsson UB, Nordlund P, Hallberg BM, X-ray structure of tRNA pseudouridine synthase TruD reveals an inserted domain with a novel fold, *FEBS Lett* **565**:59–64, 2004.
2. Bachellerie JP, Cavaillé J, Hüttenhofer A, The expanding snoRNA world, *Biochimie* **84**:775–790, 2002.
3. Cherry JM, Hong EL, Amundsen C *et al.*, Saccharomyces Genome Database: The genomics resource of budding yeast, *Nucleic Acids Research* **40**:D700–D705, 2012.
4. Hunter S, Jones P, Mitchell A *et al.*, InterPro in 2011: New developments in the family and domain prediction database, *Nucleic Acids Research* **40**:D306–D312, 2012.
5. Behm-Ansmant I, Massenet S, Immel F *et al.*, A previously unidentified activity of yeast and mouse RNA: Pseudouridine synthases 1 (Pus1p) on tRNAs, *RNA* **12**:1583–1593, 2006.

6. Behm-Ansmant I, Grosjean H, Massenet S, Motorin Y, Branlant C, Pseudouridylation at position 32 of mitochondrial and cytoplasmic tRNAs requires two distinct enzymes in *Saccharomyces cerevisiae*, *J Biol Chem* **279**:52998–53006, 2004.
7. Henras AK, Capeyrou R, Henry Y, Caizergues-Ferrer M, Cbf5p, the putative pseudouridine synthase of H/ACA-type snoRNPs, can form a complex with Gar1p and Nop10p in absence of Nhp2p and box H/ACA snoRNAs, *RNA* **10**:1704–1712, 2004.
8. King RD, Rowland J, Oliver SG *et al.*, The automation of science, *Science* **324**:85–89, 2009.
9. Sparkes A, King RD, Aubrey W *et al.*, An integrated laboratory robotic system for autonomous discovery of gene function, *J Assoc Lab Automat* **15**:33–40, 2010.
10. Fox, J, Weisberg S, *An R Companion to Applied Regression*, 2nd edn. Sage Publications, New York, 2011.
11. Meyer P, Bi-logistic growth, *Technol Forecasting Social Change* **47**:89–102, 1994.
12. DeRisi JL, Iyer VR, Brown PO, Exploring the metabolic and genetic control of gene expression on a genomic scale, *Science* **278**:680–686, 1997.
13. Brauer MJ, Saldanha AJ, Dolinski K *et al.*, Homeostatic adjustment and metabolic remodeling in glucose-limited yeast cultures, *Mole Biol Cell* **16**:2503–2517, 2005.
14. Wu G, Xiao M, Yang C *et al.*, U2 snRNA is inducibly pseudouridylated at novel sites by Pus7p and snR81 RNP, *European Mole Biol Organ J* **30**:79–89, 2011.
15. Anderson MZ, Brewer J, Singh U *et al.*, A pseudouridine synthase homologue is critical to cellular differentiation in toxoplasma gondii, *Eukaryotic Cell* **8**:398–409, 2009.
16. Patton JR, Bykhovskaya Y, Mengesha E *et al.*, Mitochondrial myopathy and sideroblastic anemia (MLASA): Missense mutation in the pseudouridine synthase 1 (PUS1) gene is associated with the loss of tRNA pseudouridylation, *J Biol Chem* **280**:19823–19828, 2005.
17. Mason PJ, Wilson DB, Bessler M, Dyskeratosis congenita — a disease of dysfunctional telomere maintenance, *Curr Mole Med* **5**:159–170, 2005.
18. Jack K, Bellodi C, Landry DM *et al.*, rRNA pseudouridylation defects affect ribosomal ligand binding and translational fidelity from yeast to human cells, *Mole Cell* **44**:660–666, 2011.
19. Zhao X, Patton JR, Ghosh SK, *et al.*, Pus3p- and Pus1p-dependent pseudouridylation of steroid receptor RNA activator controls a functional switch that regulates nuclear receptor signaling, *Mole Endocrinol* **21**:686–699, 2007.
20. Näär AM, Thakur JK, Nuclear receptor-like transcription factors in fungi, *Genes Develop* **23**:419–432, 2009.
21. Zebarjadian Y, King T, Fournier MJ, Clarke L, Carbon J, Point mutations in yeast *CBF5* can abolish *in vivo* pseudouridylation of rRNA, *Mole Cell Biol* **19**:7461–7472, 1999.
22. The Gene Ontology Consortium, Gene Ontology, *Gene Express* **25**:25–29, 2000.

Ross D. King is Professor of Machine Intelligence at the University of Manchester, UK. His research interests are in the interface between computer science and biology/chemistry. He received his B.Sc. in Microbiology in 1983 from the University of Aberdeen, M.Sc. in Computer Science from the University of Newcastle Upon Tyne in 1985, and Ph.D. in Computer Science from the Turing Institute in 1988. Since 2012, he has worked in the Manchester Institute of Biotechnology.

Chuan Lu obtained her Ph.D. from the Department of Electrical Engineering, K.U. Leuven, Belgium in 2005. She is currently a lecturer at the Department of Computer Science, Aberystwyth University. Her research interests are in the areas of bioinformatics, systems biology, clinical decision support, machine learning, and applied statistics.